The objective of this document is to develop a list of historical data requirements from the current suite of software applications in order to meet the needs of the RFC's. It will describe the current functionality of the Historical data browser and the Legacy software applications. At the end, there is a list of current differences and concerns with the web-based HDB.

1-Historical Data Browser

The Historical Data Browser is an interactive application with the Shared Windows Server. The purpose of this application is to aide users producing time series data used for model calibration.

The Historical data browser can locate a station or a group of stations based on a set of specified criteria. It performs a database query against the Historical Informix database "histdata" to get information about the Historical data stored in HNHS archives. Once the stations that meet the criteria are found they are plotted on a map as filed of circles. By selecting a particular station from the map display, the user is able to view a graphical description of data coverage, both chronologically and quantitatively. In addition, the user can compare the information for a particular data type from up to three stations, reuse queries from session to session, check the data consistency using a variety of tools, get information about the completeness of the data. All this information will allow the users to narrow the set of stations for which they wish to calculate time series data.

1-1 Data involved:

The Historical Data Browser extracts data from NHDS historical data archives and stores it on disk in a location that can be specified by the user. The actual data is generally located remotely in NHDS servers and it is retrieved via ftp. In all cases, when the data is viewed and when the time series are written, the information is stored on the user's local system. Additional disk space is not consumed during the search process. It is consumed when the data is viewed or when time series are written.

The data sources are:

- TD3200 - Daily COOP Station Data.
- TD3240 - Hourly COOP Precipitation Data.
- TD3280 - Surface Airways Observations (SAO).
- Discharge and PeakFlow – USGS.
- SNOTEL.

The data types are:

- Dew Point Temperature, Hourly.(DPTP)
- Dry Bulb (Air) Temperature, Hourly.(TMPD)
- Intermittent Peakflow Values. (PEAK)
- Pan Evaporation. (EVAP,EPAN)

- Precipitation, Daily. (PRCP,PTPX)
- Precipitation, Hourly. (PP01,PP01)
- Snow Depth at Observation Time, Point.(SNWD,SNOG)
- Snowfall.(SNOW,SNOF)
- Stream Discharge. (QME)
- Temperature, Average. (TAVG)
- Temperature, Maximum. (TMAX)
- Temperature, Minimum. (TMIN,TAMN)
- Temperature at Observation Time.(TOBS)
- Temperature of Water in Evaporation Pan, Maximum.(MXPN,TPMX)
- Temperature of Water in Evaporation Pan, Minimum.(MNPN,TPMN)
- Total Sky Coverage, Hourly.(TSKC)
- Water Equivalent of Snow Depth, Daily.(SNWE)
- Water Equivalent of Snow Depth, Point.(WTEQ,SNWE)
- Weather Summary, Daily.(DYSW)
- Wind Direction, Hourly.(WD16)
- 24-Hour Wind Movement.(WDMV,UDIS)

The USGS data that is currently not displayed is:

- Pool Elevation.(PELV)
- Reservoir Storage.(RSTO)
- Stage.(STG)

1-2 Station search:

The users need to specify a set of stations for which time series will be created. Therefore, they will need to supply some selection criteria to get the stations they want. The following is a list of criteria:

- Latitude and longitude coordinates of the area or the state where the stations are located.
- Start and End dates for data collected.
- Minimum number of years within the period of record for which data was recorded.
- Data type.
- Data Source.
- Station number.
- Station name.
- Station elevation.

The selected search criteria can be saved as a file for later or immediate use.

1-3 Map display of the stations:

All the stations that meet the supplied information by the users will be displayed on a map by the Shared Windows Server screen. The data is plotted on the map as a field of circles. A variety of information about the stations can be displayed on the map such as:

- Station's color based on observation time (default) or data source.
- Station's number according to its data source.
- Station's elevation.
- Station's moves, only for NCDC Coop stations.
- Years where data was collected for each station.

Each of the following items plotted in the map or graph has a color associated with it to highlight and draw attention to items of importance:

- Data sources. (USGS, NCDC…)
- Data types. (hourly data, daily data, hourly and daily data)
- Data values.
- Accumulated data.
- Trace data.
- Estimated data.
- Missing data.
- Basin boundaries.
- Counties.
- Rivers.

1-4 Actions performed on the map screen:

- Change the map dimension.
- Zoom in and out.
- Turn on/off rivers, basin boundaries, counties, station numbers, station move, station elevation, and years where data is collected.
- Select a station to display information about it.
- Display the data completeness with a graph representing the percent of the number of observations. For NCDC SAO data, there is no information on data completeness.
- Display the time range and elevation of the selected data.
- Display data values for each data type for a given station. The data is transferred via ftp from NHDS archives to the local machine; it can be stored locally or erased upon exit from the historical data browser.
- Change the data type being displayed or the number of data types that were selected by the user.
- Limit the date range of the data being displayed.
- Add one or group of stations to the selected station list from the map screen.
- Display station's information for a given station in the map such as name, source, latitude, longitude, and elevation.
- Compare the station's information to other stations.
- Add and remove stations from the screen.
- Reload a previously saved selected station list.
- Ability to overlay lot of information.
- Change the scale of the plot axes.
- Create time series for the currently displayed station or any station that is in the list.

- Create time series and header card file for PXPP program
- Create time series and header card file for MAT program.

1-5 Writing Time Series:

The user can create time series for any of the stations on the selected station list. Time Series are written in OH Datacard format either for one station at a time or group of stations all at once. The time series files will be written to a directory specified by the user.
After a selection of one station or group of stations is made, the station' number and name are displayed. The following items are also displayed for selection:
- Starting and ending date range for the station's collected data**.**
- The station's data types.
- Units.(English or Metric)
- Time boundaries. (Calendar year or Water year)

Since NCD has errors with accumulated data, there are two options that he user may or may not choose to review time series against NCDC data which are:
- Print NCDC values with associated flags.
- Print all NCDC accumulated data access for time series.

1-5-1 creating input for MAT program:

The user can create header card file and associated time series that are required for MAT program. The header file contains the F, G, H and Q cards. The units, time boundaries (Calendar year or water year), time range (starting year and ending year), data type(MAXT and MINT) are also selected by the user.

1-5-2 creating input for PXPP program:

The user can create header card file and associated time series that are required for PXPP program. The station number, observation time, units (English or metric), time boundaries (Calendar year or water year), time range (starting year and ending year), and the card file name need to be selected by the user. The card files will be created along with the time series selected in separate files and in a directory selected by the user.

2-Web Legacy programs

The Legacy FORTRAN programs accessed through the web were originally written for mainframe and then ported to UNIX system with a web interface put on them.
The followings are the legacy programs that are currently in use in NHDS web site:

- DLYTRAN program is used to create OH Datacard time series from daily precipitation data.
- HLYTRAN program is used to create OH Datacard time series from hourly precipitation data.

- SYNTRAN program is used to compute Point Potential Evapotranspiration (PTPE) from synoptic observations on the Surface Airways data.
- STAINV program is used to help selecting the stations to use for historical data analysis.

The web interface of each program access the historical data that is archived in NHDS inventories. The inventory of historical data consists of:

- TD3200 - Daily COOP Station Data.
- TD3240 - Hourly COOP Precipitation Data.
- TD3280 - Surface Airways Observations.

The user selects the necessary set of criteria to create OH Datacard time series or to display statistical information about the data. The input selection is entered into the web forms and it is used as input files to the Legacy programs.

2-1 Daily Observation Data Transfer Program (DLYTRAN)

The purpose of this program is to create OH Datacards time series from NCDC DSI-3200 surface land daily cooperative summary of the day.
In order to create time series, the user supplies the following information in the web form:

1- Station's number: if the number of stations is less than six stations, then the stations's ID are entered directly in the web form, otherwise the stations's ID is listed in a file (up to 900 stations). The file can be named according to the user choice and transferred to NHDS server via ftp. Then, the file name is entered in the web form. All the stations that are in the list need to be from the same state or the program will not run properly.

2- Data types: the user can choose for the following list:
- PTPX / PRCP – precipitation.
- EPAN / EVAP - pan evaporation.
- SNOF / SNOW – snowfall.
- SNOG / SNWD - point snow depth.
- SNWE / WTEQ - point snow cover water-equivalent.
- TAMX / TMAX - maximum air temp.
- TAMN / TMIN - minimum air temp.
- TPMX / MXPN - maximum pan water temp.
- TPMN / MNPN - minimum pan water temp.
- UDIS / WDMV - point wind travel.

3- Start and End dates for data collected.

4- Year type: data display either using calendar year boundaries or water year boundaries.

5- Units: data display in English or Metric units.

6- Output format of the times series can be displayed in different ways according to the user selection. The following will describe the different output format:

1- Output of the data is in a table for each station, each data type, and each year of the time requested. The table contains time series for daily precipitation with total precipitation for each month. The table's header contains the state number and name, station number and name, latitude, longitude, elevation, and the period of record requested.

2- Output of the data is in a summary table combining time for each station-data type requested. The results are calculated using the bounds of the time requested. The table's header contains the state number and name, station number and name, latitude, longitude, elevation, and the period of record requested.

3- Creation of OH Datacards time series with or without the following options:

- Display of OH Datacards with data files with missing symbols if the period of the record in the archive is shorter than the requested period of record.
- Display of OH Datacards with PXPP Station information headers. The PXPP header info will not be generated unless only PRCP data type is selected and the user requests the entire time series. The program creates the PXPP as it is getting the data.
- Display of OH Datacards with a print of NCDC data flags for data values with associated flags.
- Display of OH Datacards with a print of NCDC accumulated data cases for the time series.

4- Creation of @F, @G and @H cards only if temperature data types are selected.

2-2 Hourly Observation Data Transfer Program (HLYTRAN)

The purpose of this program is to create OH Datacard time series from NCDC DSI-3240 hourly precipitation observations. The data is retrieved from NHDS archive. The only data type involved is PTPX.

The user needs to supply the following selection criteria described bellow to create time series:

- Station number. (same as described for TD3200 above)
- Start and End dates for data collected.
- Year type: data display either using calendar year boundaries or water year boundaries.
- Units: data display in English or Metric units.
- Format of the return output. The output of the times series can be displayed in different ways according to the user selection:

1- Output of the data is in a table for each station and each month of the time requested. The header containing the state number and name, station number and name, latitude, longitude, elevation, and the period of record requested is also displayed.
2- Output of the data is in a summary table combining time for each station-data type requested and the monthly average for all the years requested. The results are calculated using the bounds of the time requested. The table's header contains the state number and name, station number and name, latitude, longitude, elevation, and the period of record requested.
3- Creation of OH Datacards time series with or without the following options:

- Display of OH Datacards with data filled with missing symbols if the period of the record in the archive is shorter than the requested period of record.
- Display of OH Datacards with PXPP Station information headers. PXPP header info will not be generated unless you request the entire time series.

2-3 Synoptic Data Transfer Program (SYNTRAN)

The purpose of this program is to create Point Potential Evapotranspiration (PTPE) data from NCDC TD-3280 Surface Airways Observations that are in NHDS archive.
In order to create PTPE time series data, the user needs to select the following criteria:

- Station number from a list of stations that is listed in the web form.
- Start and End dates for data collected.
- Year type: data display either using calendar year boundaries or water year boundaries.
- Units: data display in English or Metric units.
- The output format of PTPE can be displayed in different ways according to the user selection:
  1- Output of the data is in a table displaying computed PTPE values per year per station. The header containing the state number and name, station number and name, latitude, longitude, B3 parameter, sky cover correction factor and the period of record requested is also displayed.
  2- Output of the data is in a summary table displaying the monthly PTPE calculated from the bounds of the time requested. The table's header contains the state number and name, station number and name, latitude, longitude, B3 parameter, sky cover correction factor and the period of record requested.
  In addition, the user has the option to display intermediate data type's values used during computation. PTPE data type doesn't need to be included in the user choice since it is always displayed with table option is chosen as an output format.
  The intermediate data types involved are:
  - Average daily wind travel.
  - Weighted sky cover.
  - Average daily dew point.

- Average daily air temperature.
- Daily computed radiation.

3- Creation of PTPE in OH Datacard (time series) format. The user has the option to create the time series with or without the following selections:

- Pad beginning and ending with missing if actual values are unavailable.
- Generation of mean daily PE cards for MAPE.

2-4 Station Inventory Program (STAINV)

The purpose of this program is to help locate NCDC and NRCS SNOTEL stations form the NHDS archive and to display information about the stations. The station history information and observation times are not available for SNOTEL data.
The user will need to select a set of information from the web form in order to get the station history information.

- Source agency: The user must choose one of the following :
  - NCDC TD-3240 cooperative hourly precipitation.
  - NCDC TD-3200 cooperative summary of the day.
  - NRCS daily SNOTEL observations (high elevation data).
  - Mix daily NCDC and NRCS data.

- Data type: Any or all the following data types may be specified:
  - PTPX / PRCP – precipitation.
  - EPAN / EVAP - pan evaporation.
  - SNOF / SNOW – snowfall.
  - SNOG / SNWD - point snow depth.
  - SNWE / WTEQ - point snow cover water-equivalent.
  - TAMX / TMAX - maximum air temp.
  - TAMN / TMIN - minimum air temp.
  - TPMX / MXPN - maximum pan water temp.
  - TPMN / MNPN - minimum pan water temp.
  - UDIS / WDMV - point wind travel
- Period of record.
- Minimum number of years within the period of record for which data must exist.
- Inclusive latitude-longitude in decimal degrees to define the area to be process, or state postal code or coop state and station number.
- Return output: The program outputs the following information for the selected stations: general station information, a plot of the station's relative location, station history information, daily station observations times and charts showing time coverage for each requested element per station (will scan the raw files).

3- Difference between Web-based and HDB

The following is a list of current differences and concerns with the web-based HDB.

1- No map display for the stations.
2- No interface to make station selection.
3- No interface to display data values of the stations selected.
4- No ability to overlay different information about the stations.
5- Can not generate time series data files for each data type and for each station; they are all combined into one large file.
6- Time series data files don't have an individual file name that includes the date of coverage and data type. The name is chosen by the user.
7- The stations can only be selected using their station's numbers. With HDB, stations can be selected by entering either one of the following: station's number, station's name, latitude and longitude rang, or state.
8- Can not select stations from different states. They all need to be from the same state; otherwise the program will not run properly.
9- PXPP headers info will not be generated unless only PRCP data type is selected.
10- PXPP headers are created within the time series data files and not separately. The user will have to cut and paste the needed information in separate files.
11- MAT input file are either created combined with their associated time series in a file or separately. But, the user will need to run the form twice to get time series data files and MAT input files separately.
12- User doesn't have direct access to the data files through AWIPS. Currently data download is done through an internet process on Personal Computers. The downloaded files must then be transferred to the AWIPS system via ftp. This is because http access is not permitted thru the AWIPS firewall. At this point, ftp access is still permitted, but it will be prohibited soon with OB5 and will have to be replaced with ssh or sftp calls.